Virology 275, 294–305 (2000)
doi:10.1006/viro.2000.0498, available online at http://www.idealibrary.com on IDEAL®

Comparative Genomics of the Late Gene Cluster from *Lactobacillus* Phages

Frank Desiere, R. David Pridmore, and Harald Brüssow[1]

Nestlé Research Center, Nestec Ltd., Vers-chez-les-Blanc, CH-1000 Lausanne 26, Switzerland

Received February 10, 2000; accepted July 2, 2000

Three prophage sequences were identified in the *Lactobacillus johnsoni* strain NCC533. Prophage Lj965 predicted a gene map very similar to those of *pac*-site *Streptococcus thermophilus* phages over its DNA packaging and head and tail morphogenesis modules. Sequence similarity linked the putative DNA packaging and head morphogenesis genes at the protein level. Prophage Lj965/*S. thermophilus* phage Sfi11/*Lactococcus lactis* phage TP901-1 on one hand and *Lactobacillus delbrueckii* phage LL-H/*Lactobacillus plantarum* phage phig1e/*Listeria monocytogenes* phage A118 on the other hand defined two sublines of structural gene clusters in *pac*-site *Siphoviridae* from low-GC Gram-positive bacteria. *Bacillus subtilis* phage SPP1 linked both sublines. The putative major head and tail proteins from Lj965 shared weak sequence similarity with phages from Gram-negative bacteria. A clearly independent line of structural genes in *Siphoviridae* from low-GC Gram-positive bacteria is defined by temperate *cos*-site phages including *Lactobacillus gasseri* phage adh, which also shared sequence similarity with phage D3 infecting a Gram-negative bacterium. A phylogenetic tree analysis demonstrated that the ClpP-like protein identified in four *cos*-site *Siphoviridae* from *Lactobacillus, Lactococcus, Streptococcus,* and *Pseudomonas* showed graded sequence relationships. The tree suggested that the ClpP-like proteins from the phages were not acquired by horizontal gene transfer from their corresponding bacterial hosts.

INTRODUCTION

Bacteriophage attack has always been a major problem in industrial fermentation, especially in the dairy industry (Peitersen, 1991). The economic impact of dairy phages led to intensive research efforts on phage isolates attacking important dairy starters like *Lactococcus lactis, Streptococcus thermophilus,* and various species of the genus *Lactobacillus* (Josephsen and Neve, 1998). These starters constitute evolutionarily related sister groups of the low GC branch of Gram-positive bacteria. The evolutionary relationships between dairy bacteriophages have not yet been investigated in detail. Until now we have concentrated our comparative analysis on different *S. thermophilus* phages (Brüssow, 1999; Brüssow *et al.,* 1998; Desiere *et al.,* 1999a,b, 1998, 1997; Lucchini *et al.,* 1999a,b,c, 1998; Neve *et al.,* 1998). An analysis of their structural gene clusters demonstrated a hierarchy of sequence relationships with sequences from other *Siphoviridae* (Lucchini *et al.,* 1998). For both *cos*-site (Desiere *et al.,* 1998, 1999a,b) and *pac*-site *S. thermophilus* phages (Lucchini *et al.,* 1998), the most closely related sequences were found in temperate *Siphoviridae* infecting *L. lactis,* followed by phages infecting the genera *Lactobacillus, Leuconostoc,* and *Bacillus* (Desiere *et al.,* 1998). The structural gene cluster from *S. thermophilus* phage Sfi21 showed, in addition, a genetic organization that resembled that of *Siphoviridae* infecting Gram-negative bacteria (e.g., phage lambda), but sequence similarities were no longer detected (Desiere *et al.,* 1999a,b). Such graded relatedness is the hallmark of any evolving system. These observations could indicate that the morphogenesis modules from specific lines of *Siphoviridae* have an evolutionary history that can be retraced by comparative sequence analysis. Notably, the degree of relatedness between phages from some lines of *Siphoviridae* appears to correlate approximately with the evolutionary distance separating their bacterial hosts.

The modular theory of phage evolution has been developed on the basis of DNA–DNA hybridization experiments with lambdoid *Escherichia coli* phages in the pre-DNA sequencing era (Botstein, 1980). An initial outline of a sequence-based theory of phage evolution has been proposed (Hendrix *et al.,* 1999), but any new theoretical framework based on phage genomics is limited by the small number of phage DNA sequences in the database. Currently, there are only about 50 completed phage genomes in the database. More phage genome sequences and genome comparisons are needed to elaborate upon a sequence-based theory. A supplementary source of phage sequences can be the bacterial genome projects. In the present report we provide a comparative genome analysis of three *Lactobacillus johnsoni* prophage sequences identified in an unfinished bacterial genome project. A comparison with published phage sequences identified two major evolutionary lines

[1] To whom correspondence and reprint requests should be addressed. E-mail: harald.bruessow@rdls.nestle.com.

0042-6822/00 $35.00

for the late gene cluster within *Siphoviridae* from *Lactobacilli.* Representatives from these two lines were also detected in other genera of low-GC Gram-positive bacteria. Both lines demonstrated weak relationships with *Siphoviridae* from Gram-negative bacteria. These observations constrain further hypotheses on the evolution of *Siphoviridae.*

RESULTS

Three prophage sequences

Screening of the sequence data covering about 70% of the ongoing *L. johnsoni* strain NCC533 genome project revealed three prophage sequences, Lj965, Lj928, and Lj771 (GenBank Accession Nos. AF195900, AF195901, and AF195902). The diagnosis of prophage sequences was based upon their overall genetic organization and similarities with phage proteins from the database (Figs. 1, 2, and 3). In two cases we were able to document an unequivocal transition from phage DNA into bacterial DNA sequences. This was observed downstream of the lysin gene of prophage Lj965 where 800 bp of noncoding DNA was followed by bacterial DNA likely to encode ClpA and ClpC proteins and a DNA-directed RNA polymerase. This was also observed downstream of the lysin gene from Lj771, which was followed by bacterial DNA with links to PemK and a methionine sulfoxide reductase. Therefore, 40 kb of likely phage sequences in NCC533 (2% of its estimated 1.8-Mb genome) represents a minimum estimate. The NCC533 strain did not spontaneously release phages nor were phages released after mitomycin C induction when tested by sensitive PCR assays (data not shown).

Lj965 prophage

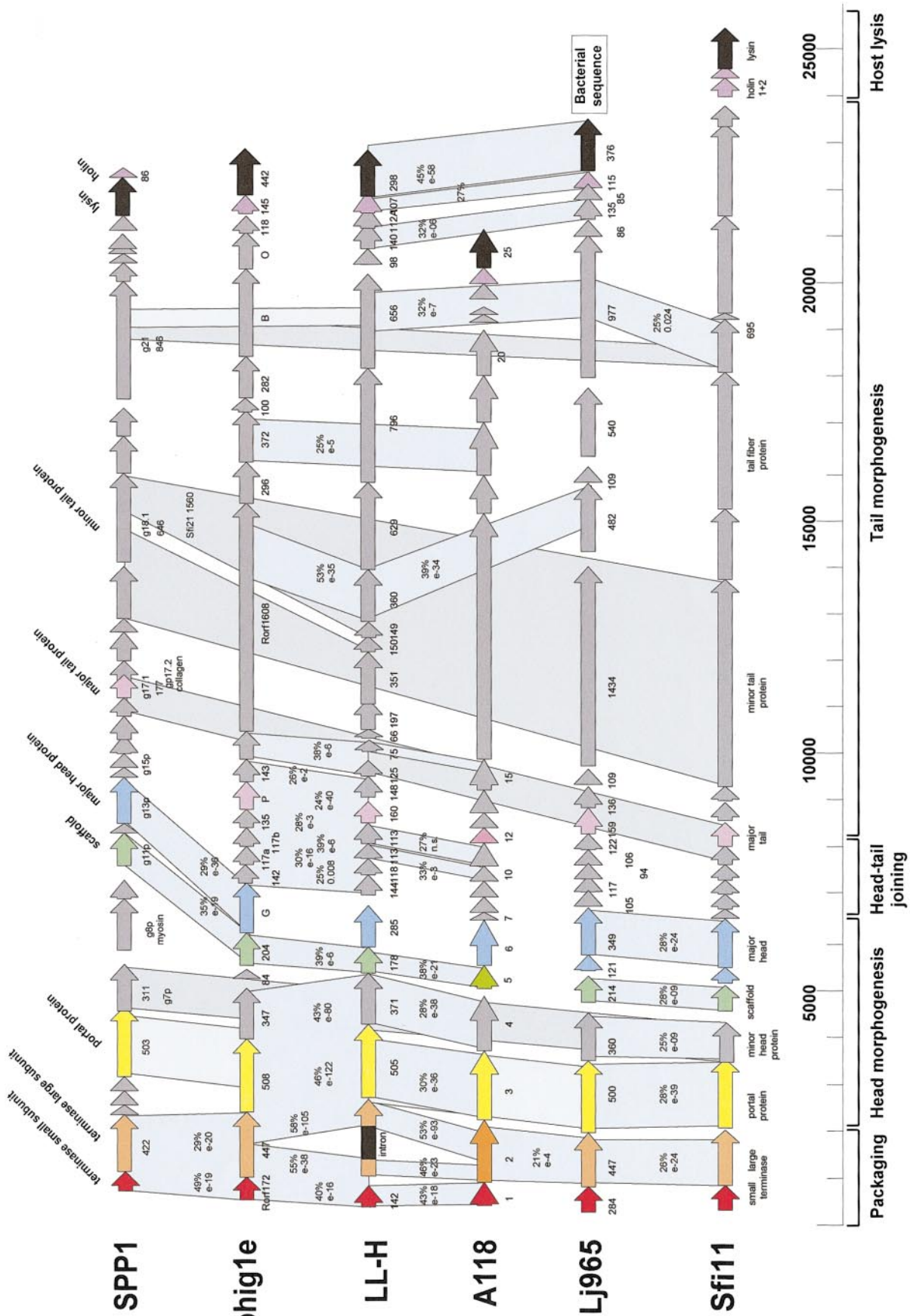
The largest identified prophage sequence was 23 kb long and comprised 25 open reading frames (orfs), all encoded on the same strand. Sequence analysis revealed that Lj965 covered the late gene cluster (DNA packaging, morphogenesis, and lysis modules) from a typical *pac*-site temperate *Siphoviridae* based on its one-to-one correspondence to similar-size genes in the DNA packaging and head and tail morphogenesis modules of *pac*-site *S. thermophilus* phages (Lucchini *et al.,* 1998, 1999b; Stanley *et al.,* 1997) and a similar correspondence to genes next to the lysis module from the *pac*-site *Lactobacillus delbrueckii* phage LL-H (Mikkonen and Alatossava, 1994) (Fig. 1). Computer-assisted analysis of the predicted proteins strongly supported the designation of a DNA packaging, head morphogenesis and lysis module, while the designation of a tail morphogenesis module was made on the basis of the similar gene map and weaker sequence similarities (Fig. 1).

The leftmost orf 284 from Lj965 predicted a protein with significant sequence similarity to a small subunit terminase from the *Bacillus subtilis* prophage PBSX, while the next orf 447 encoded a protein which matched the likely large subunit terminase of *pac*-site *S. thermophilus* phages (Table 1). Orf 500 gp resembled likely portal proteins from *pac*-site *S. thermophilus* and *B. subtilis* phages (Table 1). The genes following predict a likely minor head protein and a possible scaffold protein, encoded at corresponding map positions in *pac*-site *S. thermophilus* phages (Table 1). The adjacent orf 349 resembled major head genes from *pac*-site *S. thermophilus* phages, *Streptomyces* phage VWB, and even *E. coli* phages (Table 1). A multiple alignment of the viral proteins from these phages infecting low- and high-GC Gram-positive and Gram-negative bacteria revealed significant sequence conservation (Fig. 4). The degree of sequence conservation reflected the evolutionary distance separating the bacterial hosts (Table 1).

Also, the diagnosis of tail genes in the Lj965 prophage was backed by sequence matches. Orf 105 gp showed 24% aa identity with the *B. subtilis* phage SPP1 head completion protein gp 15 located at a corresponding map position (Fig. 1). Orf 159, which occupied the map position of the major tail gene, encoded a protein which shared 23% aa identity with the major tail proteins from coliphages HK97 and N15. In contrast, orf 1433 showed strong similarity with the putative tail measure protein from *Lactobacillus* phage adh, while orf 482 and orf 977 gps gave significant matches with minor tail proteins from *Lactobacillus* phage LL-H (Table 1) encoded at corresponding map positions (Fig. 1).

The similarity with the genome organization of *Lactobacillus* phage LL-H continued into the lysis module as demonstrated by a similar gene map (Fig. 1), significant sequence matches (Table 1), and identical transmembrane predictions of orf 117 with the LL-H holin (data not shown) despite only 27% aa identity.

FIG. 1. Comparison of the DNA packaging, head, tail, tail fiber, and lysis genes in the *Lactobacillus johnsoni* prophage Lj965 with the corresponding genes in *pac*-site phages from *Streptococcus thermophilus* (Sfi11), *L. delbrueckii* (LL-H), *L. plantarum* (phig1e), and *Bacillus subtilis* (SPP1). *Listeria* phage A118 is given as a further reference (Loessner *et al.,* 2000). Open reading frames are indicated by arrows. The number below the arrow gives the lengths of the deduced proteins in aa or the name attributed in the original publication. Corresponding genes in the five phages are indicated with the same color code; unknown genes are indicated with gray arrows. Amino acid sequence identity between phage proteins is indicated with their percentage and BLAST *E* value (expressed as a decadic logarithmic exponent). Light blue shading links related proteins. Gray shading indicate similarities between proteins of bacteriophages which are not depicted next to each other. The phage Sfi11 genes are marked with putative functions deduced from previous analysis (Lucchini *et al.,* 1999b).


SPP1
phig1e
LL-H
A118
Lj965
Sfi11
terminase small subunit
terminase large subunit
portal protein
scaffold
major head protein
major tail protein
minor tail protein
lysin
holin
Bacterial sequence
Packaging
Head morphogenesis
Head-tail joining
Tail morphogenesis
Host lysis
5000
10000
15000
20000
25000

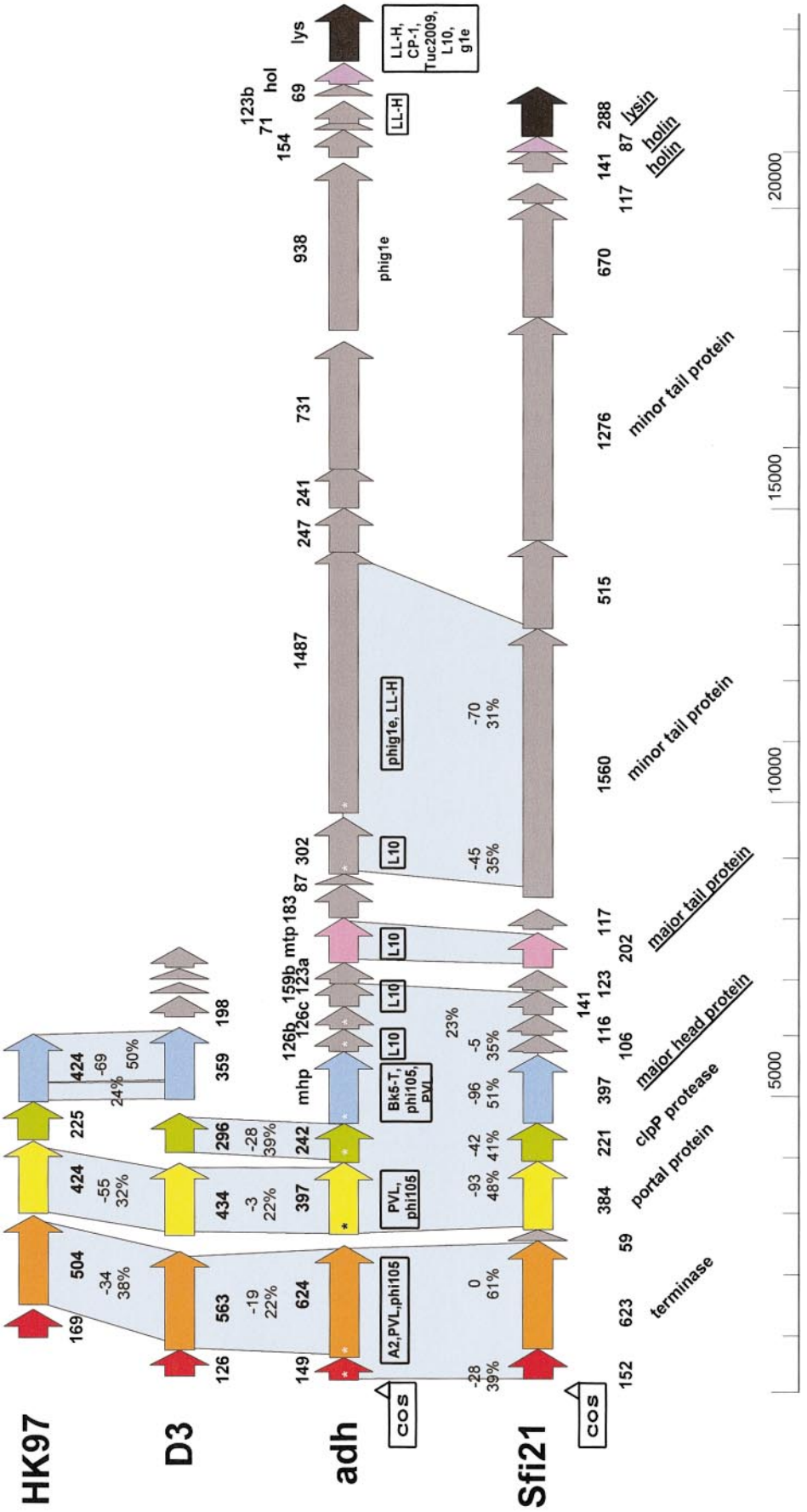


FIG. 5. Comparison of the DNA packaging, head, tail, tail fiber, and lysis genes in the *cos*-site *Lactobacillus casei* phage adh with the corresponding genes in the *cos*-site *S. thermophilus* phage Sfi21. The partial gene maps from the *Pseudomonas aeruginosa* phage D3 and the *E. coli* phage HK97 (Accession No. AF069529) are provided for comparison. The numbers next to the arrows give the lengths of the deduced proteins in aa. Some phage Sfi21 genes are identified by their putative function, functions verified by biological experiments are underlined. Related phage proteins are linked by light blue shading and the aa identity in percentages and the BLAST *E* value (decadic exponent) are given. Similarities of the adh proteins with proteins from other bacteriophages are indicated in boxes below the respective genes (key for hosts of the indicated phages- A2: *L. casei;* BK5-T: *Lactococcus lactis;* Cp-1: *Streptococcus pneumoniae;* L10: *Leuconostoc oenus;* phi105: *Bacillus subtilis;* PVL: *Staphylococcus aureus;* Tuc2009: *Lactococcus lactis*). Orfs with an asterisk indicate the best database match with Sfi21.

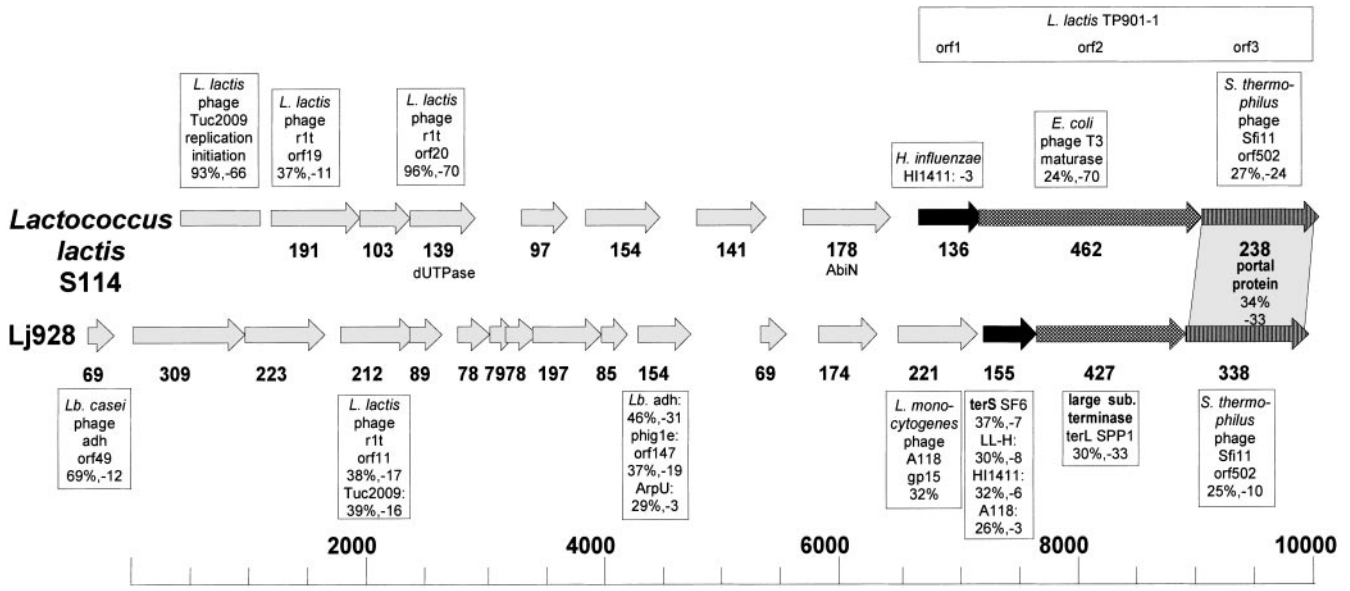


FIG. 2. Comparison of putative *Lactobacillus johnsoni* prophage Lj928 sequence with a putative prophage sequence in *Lactococcus lactis* subsp. *cremoris S114* (Accession No. Y11901). Open reading frames are indicated by a shaded arrow. The number below the arrow gives the lengths of the deduced proteins in aa. Database similarities are given next to the orfs, with their aa sequence identity in percentages and their Blast *E* value expressed as a decadic logarithmic exponent. The related portal protein genes are linked by gray shading.

Lj928 and Lj771 prophages

The 10-kb-long Lj928 prophage sequence resembled, in its genetic organization, an uncharacterized prophage sequence in *L. lactis* S114 (Prevots *et al.,* 1998) with genes showing matches to DNA replication and DNA packaging proteins (Fig. 2). Both prophage sequences showed small and large subunit terminase genes with matches to phages from Gram-positive and Gram-negative bacteria. The putative *Lactococcus* prophage showed very close matches to *L. lactis* phage TP901-1 (F. Vogensen, personal communication). The terminase gene was followed in both Lj928 and S114 prophages by a portal gene with sequence similarity to *pac*-site *L. lactis* and *S. thermophilus* phages. The 7 kb preceding the prophage DNA packaging genes showed several sequence links with *Lactococcus* and *Lactobacillus* phages over the DNA replication region. The 7-kb-long Lj771 prophage sequence shared sequence similarities with the incomplete *B. subtilis* PBSX prohage and the *Lactobacillus gasseri* phage adh over possible tail fiber genes and the likely lysis module (Fig. 3).

Comparative sequence analysis of *pac*-site *Lactobacillus* phages

The sequences from the late gene cluster of the *pac*-site *L. delbrueckii* phage LL-H (Mikkonen and Alatossava, 1994) and the *Lactobacillus plantarum* phage phig1e (Kodaira *et al.,* 1997) were retrieved from the database. Both phages shared related DNA packaging and head and tail morphogenesis modules except for short insertion/deletions (intron in LL-H, orf 84 in phig1e) and the substitution of the major head gene (Fig. 1). Interestingly, the major head protein from phage phig1e shared sequence similarity with the major head protein from the *pac*-site *B. subtilis* phage SPP1 (Becker *et al.,* 1997), but not the corresponding phage LL-H protein (Fig. 1). For the remainder of the structural genes, similarity between the two phages was limited to a C-terminal domain in the tail tape measure protein from phig1e which was also found in the corresponding protein from *cos*-site *S. thermophilus* phage Sfi21 (Desiere *et al.,* 1998). These two *pac*-site *Lactobacillus* phages also shared sequence similarity with the *Listeria monocytogenes* phage A118 over the DNA packaging and head morphogenesis modules (Fig. 1) (Loessner *et al.,* 2000). Again, the major head protein from A118 was unrelated to the corresponding phig1e or LL-H proteins, while it showed about 20% aa identity with the major head proteins from phages infecting high-GC Gram-positive bacteria (mycobacteriophage L5, *Streptomyces* phage phiC-31 (Smith *et al.,* 1999)).

With respect to the DNA packaging and head and tail morphogenesis genes, the *L. johnsoni* prophage Lj965 appears to belong to a separate evolutionary line of *Lactobacillus* phages since its closest relatives were *pac*-site *S. thermophilus* phage Sfi11 (Lucchini *et al.,* 1999b) and the *pac*-site *L. lactis* phage TP901-1 (Johnsen *et al.,* 1996; F. Vogensen, personal communication). However, the *pac*-site *B. subtilis* phage SPP1 is a linker between the Lj965/Sfi11/TP901-1 and the phig1e/LL-H/A118 lines, since it alternatively shares small groups of two adjacent genes with either group (Fig. 1).

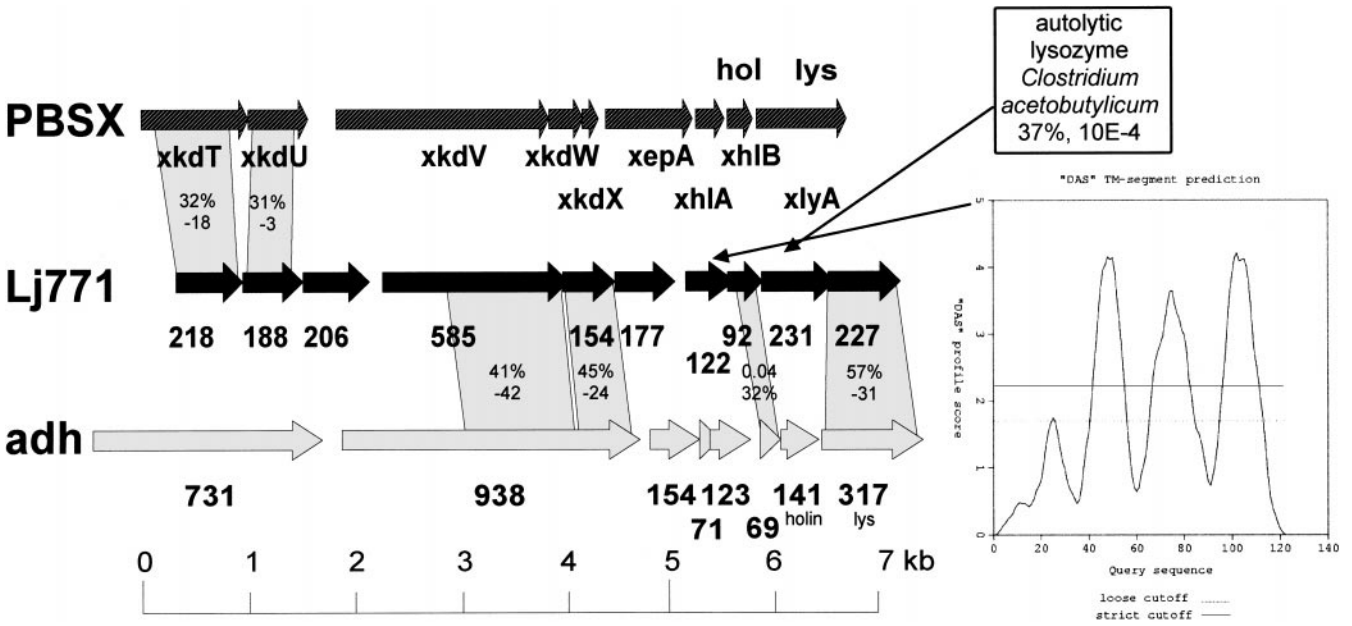


FIG. 3. Comparison of the *Lactobacillus johnsoni* prophage sequence Lj771 with the corresponding genes in the *Bacillus subtilis* prophage PBSX and a corresponding genome region in *L. gasseri* phage adh. In PBSX the name below the arrow gives the name attributed in Krogh *et al.* (Krogh *et al.,* 1996). In Lj771 and adh open reading frames are indicated by their codon lengths. Genes predicting related proteins are connected by gray shading. Similarities are given with their aa sequence identity in percentages and their *P* value (expressed as a decadic exponent). The insert provides a prediction of transmembrane segments for orf 122 from Lj771 suggesting a holin function.

Comparative sequence analysis of *cos*-site *Lactobacillus* phages

Sequence data on the late gene cluster form *cos*-site *Lactobacillus* phages are scarce: only *L. gasseri* phage adh has been sequenced entirely (Altermann *et al.,* 1999) and fragmentary data are available for the *Lactobacillus casei* phage A2 (Garcia *et al.,* 1999). Only four genes of the late gene cluster from phage adh showed sequence similarity with *pac*-site *Lactobacillus* phages. These were the tail tape measure and a likely tail fiber gene, an unattributed gene of the lysis module, and the lysin gene itself (Fig. 5). In contrast, extensive sequence similarity was shared with DNA packaging and head and tail genes from the *cos*-site *S. thermophilus* phage Sfi21 (Fig. 5). Over this region the best database matches were nearly always with phage Sfi21. However, sequence similarity between Sfi21 and adh was limited to the protein level. Seven predicted adh proteins shared sequence similarity with still other phages from low-GC Gram-positive bacteria (*Staphylococcus, Oenococcus, Bacillus*) (Fig. 5).

Interestingly, the putative terminase, portal protein, and ClpP protease from phage adh showed significant sequence similarity with phage D3 (Gilakjan and Kropinski, 1999) infecting *Pseudomonas aeruginosa,* a Gram-negative bacterium. Five of the nine phage D3 showed also significant sequence similarity with lambdoid coliphages (Fig. 5). A peculiar situation is given for the putative prohead protease and major head gene. The putative protease of phage D3 is related to that of phage adh, but not to that of coliphage HK97. In contrast, the major head protein of phage D3 shared sequence similarity with the corresponding protein from coliphage HK97 but not phage adh. Notably, the N-terminal part of the proteolytically processed prohead protein from HK97 (Duda *et al.,* 1995) showed only weak similarity with the D3 protein, while the C-terminal part representing the mature head proteins from phages D3 and HK97 shared strong sequence similarity (Fig. 4) (Gilakjan and Kropinski, 1999).

ClpP-like proteins have now been identified in four *Siphoviridae.* A phylogenetic tree analysis demonstrated that the proteins from the *S. thermophilus phage* Sfi21 (Desiere *et al.,* 1999a,b), *L. lactis* phage BK5-T (Boyce *et al.,* 1995), *L. gasseri* phage adh, and *P. aeruginosa* phage D3 belong to a single branch (Fig. 6). The cellular ClpP-like proteins from low-GC Gram-positive bacteria including *L. lactis* and close relatives of *S. thermophilus* (*Streptococcus salivarius*) and *L. gasseri* (*L. johnsoni*) belong to a distinct branch. Their closest relatives were not the ClpP-like proteins from their respective phages, but corresponding proteins from Gram-negative bacteria.

DISCUSSION

The basic tenets of the modular theory of phage evolution (Botstein, 1980; Casjens *et al.,* 1992) have withstood the test of time. However, in the era of large-scale

TABLE 1

Database Matches for the Predicted Proteins from the Putative Prophage Genes from *L. johnsoni* Lj965

Orf	Similarity	Identical aa/overaligned aa (%)	*E* value
284	*B. subtilis* prophage PBSX, small subunit terminase	62/229 (27)	10^{-15}
	B. subtilis, xre region	48/150 (32)	10^{-14}
447	*S. thermophilus* phage O1205, ORF 26 gp, large subunit terminase	104/395 (26)	10^{-24}
	Terminase large subunit *Listeria* phage A118	81/371 (21)	10^{-4}
500	*S. thermophilus* phage Sfi11, Orf 502 gp	127/446 (28)	10^{-39}
	S. thermophilus phage O1205, ORF 27 gp	123/446 (27)	10^{-38}
	Lactococcus lactis (Y11901), putative prophage protein	70/259 (27)	10^{-15}
	B. subtilis phage SPP1 portal protein	106/480 (22)	10^{-8}
360	*S. thermophilus* phage Sfi11, Orf 284 gp	39/151 (25)	10^{-8}
	S. thermophilus phage O1205, ORF 28 gp	75/336 (22)	10^{-8}
214	*S. thermophilus* phage O1205, ORF 29 gp	49/169 (28)	10^{-9}
	S. thermophilus phage Sfi11, Orf 193 gp	48/168 (28)	10^{-9}
349	*S. thermophilus* phage Sfi11, Orf 348 gp	88/311 (28)	10^{-23}
	S. thermophilus phage O1205 gp, ORF 31, major structural protein	87/311 (27)	10^{-23}
	Streptomyces phage VWB, head protein	72/305 (23)	10^{-6}
	E. coli P21	66/281 (23)	0.05
	E. coli N15 major head protein	65/288 (22)	0.09
159	*E. coli* phage HK97, major tail protein	28/120 (23)	0.04
1434	*Lactobacillus* bacteriophage phi adh orf 1487	114/379 (30)	10^{-23}
482	*S. thermophilus* phage Sfi21, Orf 1560 gp	92/207 (44)	10^{-38}
	L. delbrueckii phage LL-H, Orf 360 gp	102/257 (39)	10^{-34}
	Lactobacillus phage phig1e, Orf 1608, minor structural protein	64/164 (39)	10^{-20}
	B. subtilis prophage skin, YqbO protein		
	B. subtilis prophage PBSX, Orf 1332 gp	84/278 (30)	10^{-15}
	Staphylococcus aureus phage PVL, orf 16 gp	77/225 (34)	10^{-14}
		78/303 (25)	10^{-11}
540	*Yersinia* prophage, λ host specific protein J	28/129 (21)	0.005
	Lactobacillus phage phig1e minor structural protein	36/126 (28)	0.08
977	*L. delbrueckii* phage LL-H, minor structural protein gp58	38/118 (32)	10^{-7}
	Chlorella virus 1, hypothetical protein	160/789 (20)	10^{-4}
86	*S. pneumoniae* phage Dp-1 orf 124 gp	27/85 (31)	—
135	*L. delbrueckii* phage LL-H, orf 140 gp	32/100 (32)	10^{-6}
	L. gasseri phage adh orf 123 gp	32/108 (29)	10^{-3}
115	*L. delbrueckii* phage LL-H holin	27/99 (27)	—
376	*L. gasseri* phage adh lysin	184/300 (61)	10^{-101}
	L. delbrueckii phage LL-H lysin	102/219 (46)	10^{-57}
	S. pneumoniae phage CP-7 lysin	74/213 (34)	10^{-25}
	Leuconostoc phage 10MC lysin	65/201 (32)	10^{-18}
	Lactococcus phage Tuc2009 lysin	63/197 (31)	10^{-17}

DNA sequencing, the theory needs an extension to adapt new genomics data. A major problem for any sequence-based theory of phage evolution is the frequent lateral gene transfers between phages. This is, however, no longer a specific problem in addressing phage phylogeny. The comparative analysis of whole bacterial genomes has also revealed substantial impacts of lateral gene transfer events and led to doubts concerning the construction of bacterial phylogenetic trees based on 16 S rRNA sequences (Doolittle, 1999). Phylogenetic tree analysis has only recently been applied to bacteriophage proteins (Bruttin *et al.,* 1997; Gilakjan and Kropinski, 1999) despite the popularity of the technique in other branches of virology. Many bacterial virologists have assumed that the modular theory of phage evolution precluded such an analysis. However, frequent lateral gene transfer events complicate the interpretation of phage relationships, but they will not preclude the analysis of the evolutionary relationships between individual phage proteins. In fact, the phylogenetic trees for phage proteins provided here and in a previous report (Bruttin *et al.,* 1997) yielded consistent biological information. In the case of the phage ClpP-like protein and the phage integrase, streptococcal and lactococcal phage proteins were the closest evolutionary relatives, followed by lactobacilli phage proteins and then proteins from other phages. In both cases the similarity extended to adjacent genes: in the case of the ClpP-like protein it defined phages with a common prohead processing mechanism, while in the case of the integrase the tree defined a branch for *Siphoviridae* sharing a common organization of their lysogeny modules (Lucchini *et al.,* 1999b). In fact,

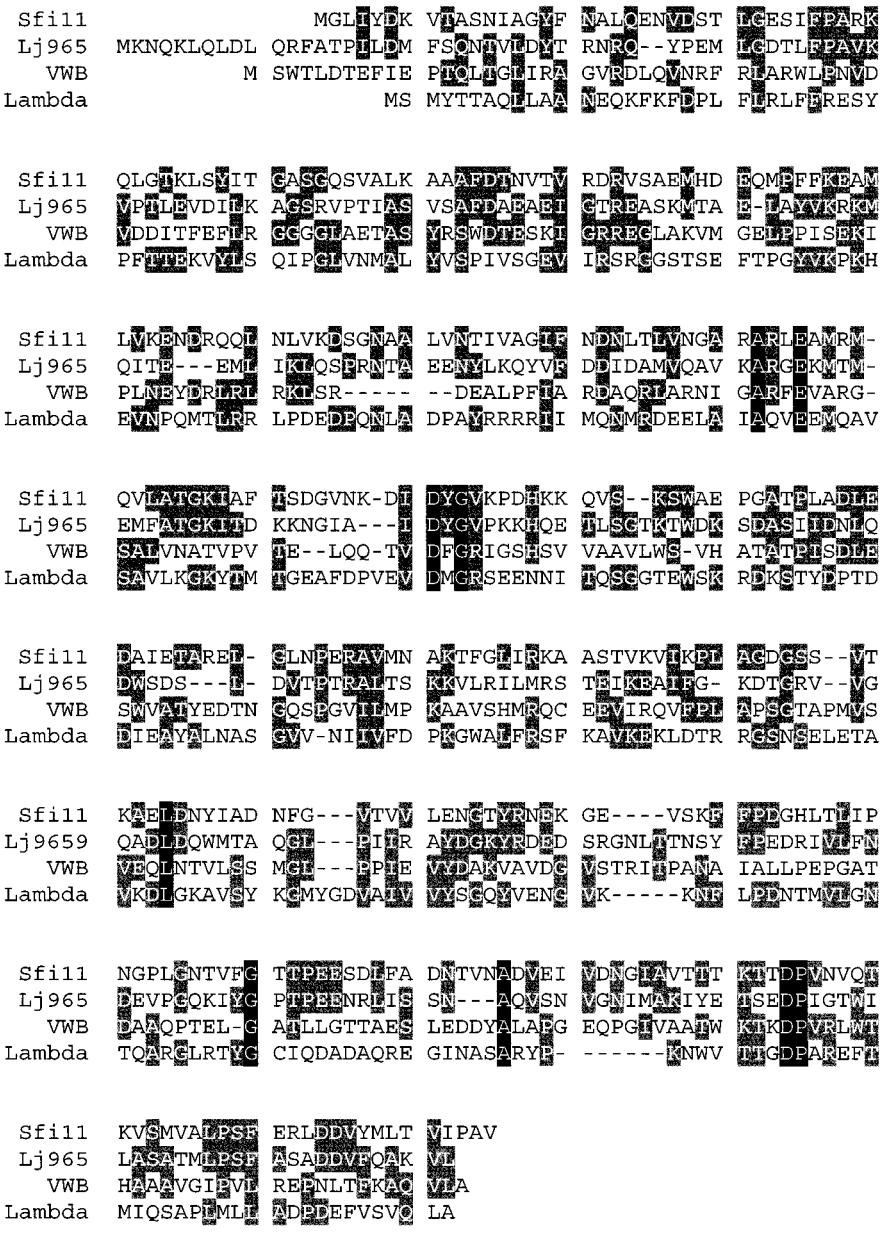


FIG. 4. Multiple alignment of the major head proteins from *S. thermophilus* phage Sfi11, *Streptomyces venezuela* phage VWB (Anne *et al.,* 1995), and *E. coli* phage λ with Lj965 orf 349 gp. Amino acid residues which are identical in at least two positions are shaded in black.

distinct branches were attributed to integrases from λ-like and L5-like temperate *Siphoviridae* (Bruttin *et al.,* 1997) which demonstrated two further distinct organizations of lysogeny related genes (Lucchini *et al.,* 1999b).

Phylogenetic tree analysis of phage proteins must, however, be interpreted cautiously. For example, in multidomain phage proteins the individual domains frequently belong to different evolutionary lineages as seen for the anti-repressor of dairy phages (Lucchini *et al.,* 1999b). Difficulties can also arise from an uncritical acquisition of sequences from the database. For example, a recent phylogenetic analysis of viral dUTPases deduced a horizontal gene transfer between lactococcal phage r1t and its *Lactococcus* host (Baldo and McClure, 1999). However, the cited *Lactococcus* host dUTPase gene is clearly part of a prophage sequence (Fig. 2). In fact, all three dUTPases identified in the *L. lactis* genome project are found within the five prophage sequences present in this dairy strain (Bolotin *et al.,* 1999).

We suggest that the evolutionary history not only of individual phage proteins, but also of entire phage modules, can be retraced by comparative sequence analysis. The phage head gene cluster is argueably the most ancient phage module and thus a suitable candidate for

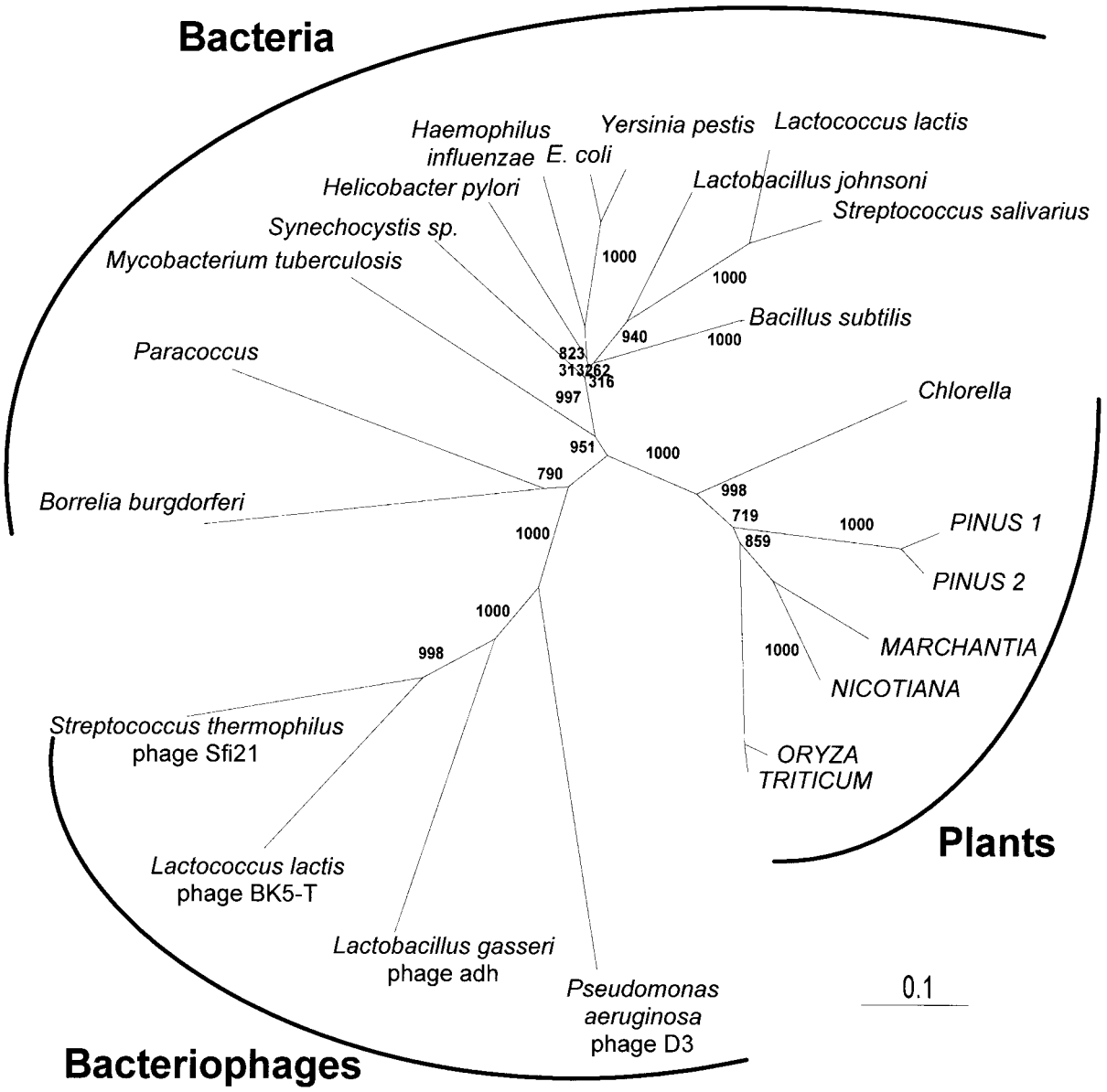


FIG. 6. Phylogenetic tree analysis of ClpP-like proteins from the indicated phages, bacteria, and plants. The unrooted tree was derived by alignment using the Clustal X program (NCBI) and subsequent representation with TreeView. The numbers at the branches give the bootstrap probabilities by the neighbor-joining method.

evolutionary analysis. Consequently, we investigated the morphogenesis module of *Lactobacillus* phages with a comparative approach. We identified two major branches of structural genes in *Lactobacillus* phages which were associated with their DNA packaging modes (*pac*-site and *cos*-site packaging). Each major branch of structural genes from *Lactobacillus* phages demonstrated close relatedness with *Siphoviridae* from evolutionarily related low-GC Gram-positive bacteria. In addition, similarity with *Siphoviridae* from Gram-negative *E. coli* were detected. This applied both to *pac*-site (major head protein) and *cos*-site (terminase, portal protein, ClpP protease) *Lactobacillus* phages (Fig. 5). *Pseudomonas* phage D3 (Gilakjan and Kropinski, 1999) is in this context an interesting link between *Siphoviridae* from Gram-negative and Gram-positive bacteria. Some D3 proteins have links only to coliphages (e.g., major head protein), one only to phages from some Gram-positive phages (ClpP-like protein); others have links to both groups of phages (e.g., large subunit terminase). Interestingly, D3 phage has 3′-extended termini, which was until now limited to phages from Gram-positive bacteria (Gilakjan and Kropinski, 1999). In addition, phage D3 demonstrated a type of head processing now identified in phages from Gram-negative and Gram-positive bacteria. Further evidence for a shared ancestry of the structural gene cluster in *Siphoviridae* from Gram-positive and Gram-negative bacteria is provided by the strikingly similar gene

map seen in *E. coli* and *S. thermophilus* (Desiere *et al.,* 1999a,b; Lucchini *et al.,* 1998). In fact, the two basic head gene constellations presented by *pac*-site and *cos*-site *Siphoviridae* from low-GC Gram-positive bacteria recalled the two basic head building concepts in lambdoid phages: that represented by coliphage λ (a scaffold protein encoded by a different gene than the major head gene) and that by coliphage HK97 (a scaffold protein proteolytically split by a phage-encoded protease from a prohead protein).

One should keep in mind three fundamental limitations of our analysis. First, when speaking of *Siphoviridae* one should realize that this term is primarily a morphological description of phages and does not represent a homogeneous group. We therefore specify that the current analysis applies to small isometric head *Siphoviridae* with a genome size of about 40 kb. Second, when speaking of the structural gene cluster or morphogenesis module one should not imagine this region as a homogenous genetic unit extending from DNA packaging to tail fiber genes. The longest cluster of linked genes which we could identify by comparative sequence analysis extended from the likely small subunit terminase to the tail tape measure gene. This cluster was observed when comparing both *pac*-site (LL-H/phig1e) and *cos*-site phages (adh/Sfi21). Several phage comparisons (e.g., Lj965/Sfi11 or LL-H/A118) defined a second smaller cluster of linked genes within this region which ranged from the terminase to the major head gene. The regions flanking the major head gene and a region within the tail tape measure gene are apparently favored modular exchange points (Sfi11/TP901-1 (Lucchini *et al.,* 1998; F. Vogensen, personal communication) secondary exchange reactions like the isolated head gene exchange in LL-H/phig1e, Fig. 1). A comparison with the phage lambda map indicated that the two end points of these linked gene clusters correspond to the borders of groups of functionally related genes (modules). A third limitation of comparative phage genomics and evolutionary analysis is the small number of phage sequences currently available in the database. Screening of bacterial genome projects for prophage sequences and their subsequent analysis might provide further insight into questions of phage evolution as demonstrated in this report. However, additional efforts are necessary. With the current ease of DNA sequence acquisition, an international consortium interested in bacteriophage evolution should direct efforts toward obtain phage sequences from a set of phages representing a wide evolutionary distribution of cultivatable bacteria.

MATERIALS AND METHODS

DNA methods

L. johnsoni chromosomal DNA was isolated as previously described (Delley *et al.,* 1990). One-microgram samples were digested with the restriction enzymes *Bam*HI, *Sac*I, *Sph*I, *Nhe*I, *Kpn*I, and *Mlu*I. These fragments were ligated into pUC19 digested with the appropriate enzymes followed by dephosphorylation and transformed into *E. coli* strain XL1-blue (Stratagene Corp.). Insert containing clones were selected on LB plates supplemented with 100 μg/ml ampicillin, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, and isopropyl-β-D-thiogalactoside. Plasmid DNA was isolated from white colonies and sequenced using the IRD-800 labeled sequencing primers (MWG) FOR (5′CTGCAAGGCG ATTAAGTTGGG3′) and REV (5′GTTGTGTGGA ATTGTGAGCGG3′) and the Thermo Sequenase kit RPN2538 (Amersham). The DNA sequence was separated and base-called using a Li-Cor Model 4000L DNA sequencer.

Sequence analysis

The Genetics Computer Group (University of Wisconsin, Madison, WI) software package was used to assemble and analyze the sequences. Open reading frames have been predicted using Clonemanager version 5.0 (Scientific & Educational Software) using ATG and GTG as possible start codons and a minimum size of 90 aa has been used. Nucleotide and predicted amino acid sequences were compared to those in the databases of GenBank release 110; EMBL [abridged], release 58; PIR-Protein release 60; SWISS-PROT release 37; PROSITE release 15. Additional database searches have been conducted using BLAST (Altschul *et al.,* 1997) and PSI-BLAST (Altschul *et al.,* 1997) at the NCBI and FASTA (Lipman and Pearson, 1985). Sequence alignments were done using the MultAlin program (Corpet, 1988), ClustalW (Thompson *et al.,* 1994), and the SIM alignment tool (Huang *et al.,* 1990). The phylogenetic tree was constructed with the Clustal X program (NCBI) and visualized with the TreeView program. Transmembrane predictions have been done using the TMpred algorithm (Hofmann and Stoffel, 1993). The Lj965, Lj928, and Lj771 prophage sequences have been submitted to GenBank under Accession Numbers AF195900, AF195901, and AF195902, respectively.

We searched the *L. johnsoni* genomic sequence dataset with bacteriophage sequences of known *Lactobacillus, Lactococcus, Streptococcus,* and *Bacillus* phages using the BLAST algorithm. As first screening probes we used bacteriophage LL-H, Accession Number M96254 (Vasala *et al.,* 1995), bacteriophage SPP1 complete nucleotide sequence, X97918 (Becker *et al.,* 1997), *B. subtilis* DNA (28-kb PBSX/skin element region) Z70177 (Krogh *et al.,* 1996), *Lactobacillus* bacteriophage phig1e complete genomic DNA, X98106 (Kodaira *et al.,* 1997), and *S. thermophilus* bacteriophage Sfi21, Sfi19, and Sfi11 complete genomes (Lucchini *et al.,* 1999b).

ACKNOWLEDGMENTS

We thank A.-C. Pittet and M.-C. Zwahlen for the elaboration of the *Lactobacillus johnsoni* prophage sequences, A. Hillier and F. Vogensen for sharing unpublished *Lactococcus lactis* bacteriophage sequences with us, S. Lucchini for valuable discussions, C. Walker for critically reading the manuscript, and the Swiss National Science Foundation for financial support of Frank Desiere in the framework of its Biotechnology Module (Grant 5002-044545/1).

REFERENCES

Altermann, E., Klein, J. R., and Henrich, B. (1999). Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage phi adh. *Gene* **236,** 333–346.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402.

Anne, J., Fiten, P., Van Mellaert, L., Joris, B., Opdenakker, G., and Eyssen, H. (1995). Analysis of the open reading frames of the main capsid proteins of actinophage VWB. *Arch. Virol.* **140,** 1033–1047.

Baldo, A. M., and McClure, M. A. (1999). Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. *J. Virol.* **73,** 7710–7721.

Becker, B., de la Fuente, N., Gassel, M., Gunther, D., Tavares, P., Lurz, R., Trautner, T. A., and Alonso, J. C. (1997). Head morphogenesis genes of the *Bacillus subtilis* bacteriophage SPP1. *J. Mol. Biol.* **268,** 822–839.

Bolotin, A., Mauger, S., Malarme, K., Ehrlich, S. D., and Sorokin, A. (1999). "Lactic Acid Bacteria: Genetics, Metabolism and Applications" (W. N. Konings, O. P. Kuipers, and J. H. J. Huis in't Veld, Eds.), pp. 27–76. Kluwer Academic, Dordrecht.

Botstein, D. (1980). A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354,** 484–490.

Boyce, J. D., Davidson, B. E., and Hillier, A. J. (1995). Sequence analysis of the *Lactococcus lactis* temperate bacteriophage BK5-T and demonstration that the phage DNA has cohesive ends. *Appl. Environ. Microbiol.* **61,** 4089–4098.

Bruttin, A., Desiere, F., Lucchini, S., Foley, S., and Brüssow, H. (1997). Characterization of the lysogeny DNA module from the temperate *Streptococcus thermophilus* bacteriophage Sfi21. *Virology* **233,** 136–148.

Brüssow, H. (1999). "Encyclopedia of Virology" (R. Webster and A. Granoff, Eds.), Second ed., pp. 1253–1262. Academic Press, London.

Brüssow, H., Bruttin, A., Desiere, F., Lucchini, S., and Foley, S. (1998). Molecular ecology and evolution of *Streptococcus thermophilus* bacteriophages—A review. *Virus Genes* **16,** 95–109.

Casjens, S., Hatfull, G. F., and Hendrix, R. (1992). Evolution of dsDNA tailed-bacteriophage genomes. *Semin. Virol.* **3,** 383–397.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16,** 10881–10890.

Delley, M., Mollet, B., and Hottinger, H. (1990). DNA probe for *Lactobacillus delbrueckii*. *Appl. Environ. Microbiol.* **56,** 1967–1970.

Desiere, F., Lucchini, F., and Brüssow, H. (1999a). Comparative sequence analysis of the DNA packaging, head and tail morphogenesis modules in the temperate *cos*-site *Streptococcus thermophilus* bacteriophage Sfi21. *Virology* **260,** 244–253.

Desiere, F., Lucchini, S., Bruttin, A., Zwahlen, M. C., and Brüssow, H. (1997). A highly conserved DNA replication module from *Streptococcus thermophilus* phages is similar in sequence and topology to a module from *Lactococcus lactis* phages. *Virology* **234,** 372–382.

Desiere, F., Lucchini, S., and Brüssow, H. (1998). Evolution of *Streptococcus thermophilus* bacteriophage genomes by modular exchanges followed by point mutations and small deletions and insertions. *Virology* **241,** 345–356.

Desiere, F., Mahanivong, C, Hillier, A. J., Davidson, B. E., and Brüssow, H. (1999b). Comparative analysis of *cos*-site *siphoviridae* from low gc content gram-positive bacteria: A genomics-based analysis of phage evolution. (submitted).

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284,** 2124–2129.

Duda, R. L., Martincic, K., Xie, Z., and Hendrix, R. W. (1995). Bacteriophage HK97 head assembly. *FEMS Microbiol. Rev.* **17,** 41–46.

Garcia, P., Ladero, V., Alonso, J. C., and Suarez, J. E. (1999). Cooperative interaction of CI protein regulates lysogeny of *Lactobacillus casei* by bacteriophage A2. *J. Virol.* **73,** 3920–3929.

Gilakjan, Z. A., and Kropinski, A. M. (1999). Cloning and analysis of the capsid morphogenesis genes of Pseudomonas aeruginosa bacteriophage D3: Another example of protein chain mail? *J. Bacteriol.* **181,** 7221–7227.

Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl. Acad. Sci. USA* **96,** 2192–2197.

Hofmann, K., and Stoffel, W. (1993). TMbase—A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **347,** 166.

Huang, X. Q., Hardison, R. C., and Miller, W. (1990). A space-efficient algorithm for local similarities. *Comput. Appl. Biosci.* **6,** 373–381.

Johnsen, M. G., Appel, K. F., Madsen, P. L., Vogensen, F. K., Hammer, K., and Arnau, J. (1996). A genomic region of lactococcal temperate bacteriophage TP901–1 encoding major virion proteins. *Virology* **218,** 306–315.

Josephsen, J., and Neve, H. (1998). "Bacteriophages and Lactic Acid Bacteria" (S. Salminen and A. Von Wright, Eds.), 2nd rev. ed., pp. 385–436. Dekker, Basel.

Kodaira, K. I., Oki, M., Kakikawa, M., Watanabe, N., Hirakawa, M., Yamada, K., and Taketo, A. (1997). Genome structure of the *Lactobacillus* temperate phage phi g1e: The whole genome sequence and the putative promoter/repressor system. *Gene* **187,** 45–53.

Krogh, S., O'Reilly, M., Nolan, N., and Devine, K. M. (1996). The phage-like element PBSX and part of the skin element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. *Microbiology* **142,** 2031–2040.

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227,** 1435–1441.

Loessner, M. J., Inman, R. B., Lauer, P., and Calendar, R. (2000). Complete nucleotide s, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes:* Implications for phage evolution. *Mol. Microbiol.* **35,** 324–340.

Lucchini, S., Desiere, F., and Brüssow, H. (1998). The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to *Siphoviridae* from a wide range of bacterial hosts. *Virology* **246,** 63–73.

Lucchini, S., Desiere, F., and Brüssow, H. (1999b). Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. *J. Virol.* **73,** 8647–8656.

Lucchini, S., Desiere, F., and Brüssow, H. (1999). Similarly organized lysogeny modules in temperate *Siphoviridae* from low GC content Gram-positive bacteria. *Virology* **263,** 427–435.

Lucchini, S., Desiere, F., and Brüssow, H. (1999a). The genetic relationship between virulent and temperate *Streptococcus thermophilus* bacteriophages: Whole genome comparison of *cos*-site phages Sfi19 and Sfi21. *Virology* **260,** 232–243.

Mikkonen, M., and Alatossava, T. (1994). Characterization of the genome region encoding structural proteins of *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H. *Gene* **151,** 53–59.

Neve, H., Zenz, K. I., Desiere, F., Koch, A., Heller, K. J., and Brüssow, H. (1998). Comparison of the lysogeny modules from the temperate *Streptococcus thermophilus* bacteriophages TP-J34 and Sfi21: Implications for the modular theory of phage evolution. *Virology* **241,** 61–72.

Peitersen, N. (1991). Practical phage control. *Bull. Int. Dairy Fed.* **263,** 1–43.

Prevots, F., Tolou, S., Delpech, B., Kaghad, M., and Daloyau, M. (1998). Nucleotide sequence and analysis of the new chromosomal abortive infection gene abiN of *Lactococcus lactis* subsp. *cremoris* S114. *FEMS Microbiol. Lett.* **159,** 331–336.

Smith, M. C., Burns, R. N., Wilson, S. E., and Gregory, M. A. (1999). The complete genome sequence of the *Streptomyces* temperate phage phiC31: Evolutionary relationships to other viruses. *Nucleic Acids Res.* **27,** 2145–2155.

Stanley, E., Fitzgerald, G. F., Le Marrec, C., Fayard, B., and van Sinderen, D. (1997). Sequence analysis and characterization of phi O1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology* **143,** 3417–3429.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22,** 4673–4680.

Vasala, A., Valkkila, M., Caldentey, J., and Alatossava, T. (1995). Genetic and biochemical characterization of the *Lactobacillus delbrueckii* subsp. *lactis* bacteriophage LL-H lysin. *Appl. Environ. Microbiol.* **61,** 4004–4011.